



# USING BALANCED ASSESSMENT SYSTEMS TO IMPROVE STUDENT LEARNING AND SCHOOL CAPACITY: AN INTRODUCTION

Brian Gong  
Center for the Improvement of Educational Assessment, Inc.  
Paper commissioned by CCSSO and the R&D Consortium sponsored by Renaissance Learning  
March 3, 2010

## Executive Summary

---

This report includes:

- ↻ A rationale for why balanced and comprehensive assessment systems are necessary.
  - ↻ Definitions of the terms balanced, comprehensive, coherent, and assessment system
  - ↻ Definition of the essential components and relationships of a coherent and balanced assessment system and criteria for determining the presence and quality of such components and relationships
  - ↻ A framework (template/model) for a balanced and coherent assessment system that represents a consensus of the states and sponsors involved in this project
  - ↻ Some examples of what a balanced and coherent assessment system could, in fact, look like. The examples reflect different approaches states may take.
  - ↻ References to help guide further learning about balanced assessment systems.
-

---

## Table of Contents

---

EXECUTIVE SUMMARY -----	I
TABLE OF CONTENTS -----	II
INTRODUCTION -----	1
How the Report is Organized-----	1
HOW CAN ASSESSMENT HELP IMPROVE STUDENT ACHIEVEMENT AND SCHOOL CAPACITY? -----	2
WHAT IS A “BALANCED ASSESSMENT SYSTEM”?-----	6
What are the characteristics of a balanced assessment system? -----	7
Creating More Balanced Summative Assessment -----	10
Why is it important for states to support balanced assessment systems? -----	11
How do purpose and use affect the design of an assessment?-----	13
What is the relation between assessment, curriculum, and instruction in a balanced assessment system?-----	16
HOW CAN STATES SUPPORT THE DEVELOPMENT AND IMPLEMENTATION OF BALANCED ASSESSMENT SYSTEMS? -----	19
DEVELOPING COMPREHENSIVE ASSESSMENT SYSTEMS -----	21
WHERE CAN I GO FOR MORE INFORMATION? -----	24

---

## List of Exhibits

---

Exhibit 1: Different types of tests for different purposes and audiences-----	10
Exhibit 2: Chronological View of Administration of Comprehensive Assessments -----	10
Exhibit 3: Two design dimensions of formative assessment, with illustrative approaches -----	11
Exhibit 4: Four views of school performance -----	12
Exhibit 5: Illustrations of information provided by summative, interim, and formative assessment-----	13, 14
Exhibit 6: Design of interim assessment measures-----	16
Exhibit 7: Student Proficiency Case Scenario -----	19
Exhibit 8: District-School-Classroom Case Scenario-----	21
Exhibit 9: Tiered Levels of Intervention -----	22

# USING BALANCED ASSESSMENT SYSTEMS TO IMPROVE STUDENT LEARNING AND SCHOOL CAPACITY: AN INTRODUCTION

Brian Gong

Center for the Improvement of Educational Assessment, Inc.

Paper commissioned by CCSSO and the R&D Consortium sponsored by Renaissance Learning

March 3, 2010

## Introduction

What is a “balanced assessment system” and why is one needed?

The simple answer to why a balanced assessment system is needed is because state assessment systems that focus on once-a-year summative testing do not provide sufficient information to improve student learning and school capacity. A balanced assessment system supplements the information provided by summative state tests to more completely inform improving learning and school decisions.

This report helps state policy makers and others understand what a balanced assessment is, why it is needed, and how one might get started building one. To help with the task of building a balanced assessment system, this report provides some examples drawn from the work going on in states, districts, and schools. It also provides a self-guiding checklist of key questions and elements that states and others can use to examine and improve their own assessment systems.

### *How the Report is Organized*

The report addresses five main questions:

1. How can assessment help improve student achievement and school capacity?

2. What is a “balanced assessment system”?

3. How do balanced assessment systems work to improve student achievement and school capacity?

4. How can states support the development and implementation of balanced assessments systems?

5. Where can I go for more information?

Each question is discussed at three levels. The first level gives the main points, often dealing with policy implications. The second level provides a more detailed view, suitable for those who want to understand a topic more deeply. The third level discusses more technical and operational aspects, relevant especially to those who wish to implement a balanced assessment program or understand a particular educational setting.

The first level always comes first in the section. It is often denoted with bulleted text. The second level is presented as the main text of the section. The third level is found in a shaded box to the bottom or side of the section. In addition to these three levels, main points are also emphasized through boxes and graphics throughout the report.

### Main Points about Balanced Assessment Systems

- Assessment systems are essential to improve student learning and school capacity because assessment provides essential data to monitor performance and inform decisions. The purpose of assessment is to inform constructive action.
- Balanced assessment systems include a variety of instruments and processes to produce information and data for informing policy makers, school administrators, parents, and teachers and students. State tests alone cannot provide the necessary information. Summative, interim, and formative assessment all have a role to play.
- Just as policies and actions must be coherent from the statehouse to the schoolhouse, so the components of a balanced assessment system must be designed to work together coherently. Coherence comes not only through the design of the assessments, but through how they are used.
- There are many ways a balanced and coherent assessment system can be developed. States can support development of such systems through a variety of means. If states do not actively support such development there is a greater likelihood that the assessment systems will remain incomplete or incoherent. Users should consider carefully the costs/benefits of using common, custom, or commercially developed assessment instruments. A key consideration should be developing expertise is how to use the assessment information at individual and institutional levels.
- There are many resources that can help inform efforts to develop a balanced assessment system.

## How can assessment help improve student achievement and school capacity?

Assessment can be a powerful support for improving student achievement and school capacity by providing credible, focused information on overall patterns of performance, strengths and weaknesses related to student learning and instructional programs, and feedback on how to improve. Policymakers, educators, and students need assessment information in order to know what is working and what needs to be changed. Assessment is crucial to move from opinions to informed action. Assessment may be more formal—using carefully designed assessment instruments—or more informal—such as learning from observations, conversations, and impressions. All such information may be helpful. However, using more systematic information helps avoid

making decisions based on inaccurate, incomplete, or outdated information.

Every state has a large-scale assessment system which has been tied to the school accountability provisions of the federal No Child Left Behind Act (NCLB). There is increasing agreement that more balanced and comprehensive assessment systems must be developed to support the educational improvement necessary to meet the United States' needs of excellence and equity for the 21st Century. However, while examples of a variety of more extensive assessment approaches exist, there is no popular consensus of what a balanced and coherent assessment system would look like, nor a discussion of the core features and rationales underlying such a system design.

Recently there has been a proliferation of “interim” assessment measures adopted by districts. Many states similarly are interested in interim and formative assessments. This framework for a balanced and coherent assessment system will help states, districts, and other entities coordinate their efforts to create and use more comprehensive assessment systems. In addition, such a framework would be helpful in developing policies, infrastructures, and research to inform the use and improvement of balanced assessment systems to support improved student learning and school capacity.

Many states are supporting the development of balanced assessment systems to promote student learning and increased school capacity. Balanced assessment systems include a range of assessment measures designed to inform understanding and action from state policy to classroom instruction. Having accurate and timely information is essential to improve all aspects of education. Much of that information comes from assessment systems. States have long had state assessment systems that measure student achievement towards state-established proficiency

levels on state content standards. It has become apparent to state policymakers, educators, and researchers that more and different types of information than have been provided by state assessments are needed to inform the desired improvements in student learning. This additional information can be provided by balanced assessment systems.

This report provides an introduction to balanced assessment systems. It explains what a balanced assessment system consists of, and defines essential terms. The report uses interim assessments as a focus to illustrate the principles that an assessment must be designed for a particular use, but that a balanced assessment system will provide comprehensive and coherent information to

support the goals of fostering increased student learning and developing school capacity. The report includes a summary of characteristics that a state, district, or other user might use to design or evaluate a balanced assessment system. The report closes with examples of how a state might go about supporting the development and implementation of balanced assessment systems.

A *BALANCED ASSESSMENT SYSTEM* helps answer the questions of:

- What happened? Strengths & weaknesses?
- Why?
- How can it be improved?

It provides this information at the levels of:

- State/National
- District
- School
- Classroom/Individual

It includes

- Summative tests
- Interim assessment measures
- Formative assessment process

It addresses all important areas.

It includes support so the assessment data may be interpreted and used appropriately.

### *A Brief Historical Context to Assessment in the U.S.*

To understand the current call for balanced and comprehensive assessment systems, one must understand the current context of assessment in the United States.

Of the many influences shaping current assessment policy and practice, three have been selected as major to this paper. These three influences in turn provide an outline of what the calls for improvement must address. These influences are: a) the centrality of standards-based reform in American education since the mid-1980's, b) the increasing role of states in defining and implementing central assessment systems associated with accountability, and c) the emerging limitations of state-level assessment and accountability systems to improve student learning and school system performance in a wide range of circumstances.

#### **Standards-Based Reform**

“Standards-based reform” refers to the school of thought that holds that to improve education, the key strategy is to implement “standards.” The elements of standards-based reform include common *content standards* that describe what students should know and be able to do and common *performance standards* that describe how well students should be able to do those things. The standards movement sought to address three major challenges. First, the gap in learning opportunities and outcomes between wealthier and poor students, majority and minority children, and higher-performing and “lower-performing students within classrooms was linked to differential expectations. Differential performance was due largely not to differences in ability, but rather differences in expectations by teachers, parents, and the students themselves. These differences in expectation were reified into differences in opportunity as students within classrooms were sorted into the “Bluebirds” and “Redbirds”; students within schools were tracked into “Honors,” “Regular,” and “Remedial” classes with very different content and curricula, and students graduated from schools with very different notions of what represented “a good education.”

One solution to poor opportunities born of low expectations was to establish common expectations of what students should learn. These learning expectations were denoted “content standards.” Having common content standards would help counteract varied and low expectations.

#### **States' Role in Establishing Assessment and Accountability Systems**

For most states, “local control” has been the tradition where school districts set curricula, graduation requirements, hired and trained teachers. Teachers have traditionally set their grading standards. In the 1970's many states set minimum competency graduation standards, often accompanied by a state test. However, these standards were low enough that they were perceived as being applicable to only a small percentage of students, and there were few school consequences to not doing well on these exit exams. However, beginning in the late 1980's states became the forefront of supporting standards-based educational reform. States had the power and resources—often aided by court cases dealing with financial equity that clarified that states were responsible for the quality of education—to establish common content and performance standards, implement state assessment systems, and enact state-wide school accountability systems. These proficiency-based systems, unlike the previous minimum-competency graduation systems, were intended to affect K-12 education and were set high enough that they did. By 2000 48 states had a state assessment and accountability system. With the passage of the federal *No Child Left Behind Act* (NCLB) in 2001, states were clearly empowered and required by the federal government to develop and implement standards-based reforms.

The NCLB legislation extended what most states had done to new areas: annual testing in grades 3-8 and at least once in high school, strong consequences including mandatory offering of public school choice and “supplemental educational services” (such as tutoring) for students in schools that did not meet the accountability performance criteria, a much more demanding—and many said unrealistic—level of performance with 100% of the students required to score proficient or above by 2013-14, extension of the assessment requirements to special populations including students with disabilities and students with limited English language proficiency, accountability based on disaggregated subgroups as well as the school as a whole meeting the performance criteria, and external technical reviews of the state systems of standards, assessments. As a result, state assessment and accountability systems have become the focal source of information when people seek to measure learning and inform instruction.

### Limitations of State-level Standards-based Reform

Many people have come to recognize the limitations of state-level assessments in providing information useful to inform instruction and school improvement. Perhaps more radically, some people believe that state-level standards-based reform itself is unable to produce the desired improvements to address equity concerns or to address new goals of excellence, such as continuous improvement for higher-performing students or a renewed call for “college readiness.”

Some people advocate for better large-scale assessments, perhaps using more performance-based formats or multiple measures. Others similarly advocate for expanded or more sophisticated measurement uses of assessment data, using, for example, growth models or value-added models of statistical analysis that reflect not only performance at a point in time, but changes in performance over time by the same students, or changes in relation to some expectation or attributable to some agent. While these types of advocated changes might make an assessment system more comprehensive, they leave the basic structure of current state assessments largely the same.

Other people argue for the involvement beyond states of districts, schools, teachers, and students in the assessment and accountability endeavor. These people focus on the need for assessments that inform instruction, programs, and policies in ways that state-level assessments do not. These types of advocated changes expand the current standards-based model from the state to include other levels in the educational system.

A third type of change is advocated by those who say that the “theory of action” underlying current standards-based reform is inadequate and should be expanded or modified. That theory of action holds that methods for reaching the goals should be left to local control because schools will invent ways to meet the goals as long as the goals are clear, accurately and timely feedback is provided by means of assessments, and proper incentives are provided through accountability. In contrast, advocates of more systemic change to the standards-based theory of action point to the wide-spread failure of NCLB and previous state systems to meet their proclaimed goals, particularly among the most disadvantaged students, lowest performing subgroups, and most challenged schools. These people advocate for more complete instructional models, stronger curricula, better teacher training, and more innovative ways to deal with structural issues such as the distribution of higher quality teachers within school districts and across less socially desirable work environments such as inner city and rural schools.

### Main Points about “What is a Balanced Assessment System”?

- ↻ A balanced assessment system includes content standards and assessments.
- ↻ Content standards describe clearly what should be learned (knowledge, skills, competencies, behaviors, habits of mind, etc.), and a level of achievement as well (are students to be able to understand the knowledge, apply it, etc.).
- ↻ Assessments are constructed to measure students’ attainment of the content standards, and report on their level of achievement in clear and useful ways.
- ↻ The standards and assessments are related coherently in a system that supports achievement of the purposes through meaningful interpretation and use.
- ↻ The system is balanced appropriately between accountability and support of learning. This typically means that there are other purposes in addition to accountability, and other assessments in addition to the state’s once-a-year summative assessment.

## What is a “balanced assessment system”?

A balanced assessment system includes the components of content standards and assessment measures. They are organized within a system that has coherence and balance. Each of these attributes is defined below.<sup>1</sup>

### *Standards*

Simply stated, standards are clear and agreed upon expectations for learning that may take the form of certain knowledge, skills, competencies, or behaviors. They usually have levels of expectation that accompany them and a sorting process for determining levels of achievement as well.

### *Assessment*

Assessment is a process for eliciting evidence of that achievement from the learner. The outcomes or data from that process can be used to immediately inform and alter the instructional process while actually engaged in the learning as with “formative assessment;” to help monitor or gauge the progress of learning and perhaps predict success as with “interim assessment;” or to “sum-up” the learning at the end of the instructional unit as with

<sup>1</sup> This section is quoted with minor edits from Philip, F. (2008). “What Do We Mean by Coherent and Balanced Assessment Systems?” Washington, DC: CCSSO.

“summative assessment.” The test or other measures used in eliciting that evidence can range from simple teacher observation of classroom behavior to formal, large-scale tests that are highly complex, sophisticated, and psychometrically sound—but offer little help to the teacher or learner because of their separation and distance from the learning process. The stakes for doing well can range from “try it again” to “dire and life changing” consequences. And finally, depending on the design and intentions of the assessment or test, the outcome data can be used for immediate information and support in the learning process, or for educational accountability purposes, evaluating the effectiveness of the instructional process.

### *System*

To be a system or a thoughtful, coordinated, and comprehensive process for assuring that the curriculum, the instructional program, and the assessment process are in fact part of a cohesive vision. Depending on the state, that may be driven by a state education agency that has centralized power, or in states where the SEA has minimal power over local districts, as a part of the Chief’s educational leadership role.



### *Coherence*

All components of the curriculum, the instructional program, and the assessment process are aligned and form a coherent set of expectations for learning shared by the state, the districts, and each teacher in the classroom along with a good understanding of the natural learning progressions as the student moves toward competency. Unlike a shared vision, this coherence must be operational and supported by the necessary new teacher training and professional in-service for teachers in the classroom.

### *Balance*

Finally, the system must be balanced, exhibiting an

appropriate weighting and distribution of learning process support and accountability needs. There is no magic number of formative measures/interim measures/or summative tests—the needs for information about the learning process determine how much each type of assessment is applied and how the data are used. However, most educators agree that the balance must reflect the absolute importance of the support for learning over the simple measurement of learning for accountability. Whenever and wherever this process takes place successfully, the role of accountability becomes transformed to a different question with positive consequences.

#### A balanced assessment system addresses these important dimensions:

- ∞ Coherently informs different actors responsible for different levels of the educational system, including state/national, district, school, classroom/individual
- ∞ Provides “vertical” information integrating summative, interim, and formative assessments
- ∞ Is comprehensive enough to inform different purposes, including accountability, program improvement, and instruction
- ∞ Inclusive of students – provides appropriate assessment for all students
- ∞ Inclusive of valued content & skills
- ∞ Provides both diagnosis and prescription information of “what is” and “what should be done” to improve

#### *What are the characteristics of a balanced assessment system?*

A balanced assessment system *consists of tools to assess and inform learning*. The assessment tools inform policy, programs, and individual teachers and learners in a coherent and coordinated manner. All important areas are included.

A balanced assessment system *typically includes at least three levels of assessment measures*: summative measures (such as annual state tests), interim assessment measures (such as common tests administered by districts or schools), and formative assessment measures (an assessment process used to direct classroom learning and teaching). These measures help provide coherent information across the levels of the educational system: state/national,

district, school, classroom, and individual student/teacher. Importantly, a balanced assessment system also includes supports to use the assessment information appropriately.

A balanced assessment system *provides the appropriate information needed by different levels of the educational system*. For example, the needs of state/national policymakers are somewhat different from district school administrators, who need to provide leadership regarding their local educational programs, resources, and staff. And the concerns of teachers, students, and parents about how to help individual students represent another level.

Exhibit 1 illustrates that different levels of the educational system are typically concerned with different types of decisions, and so need different

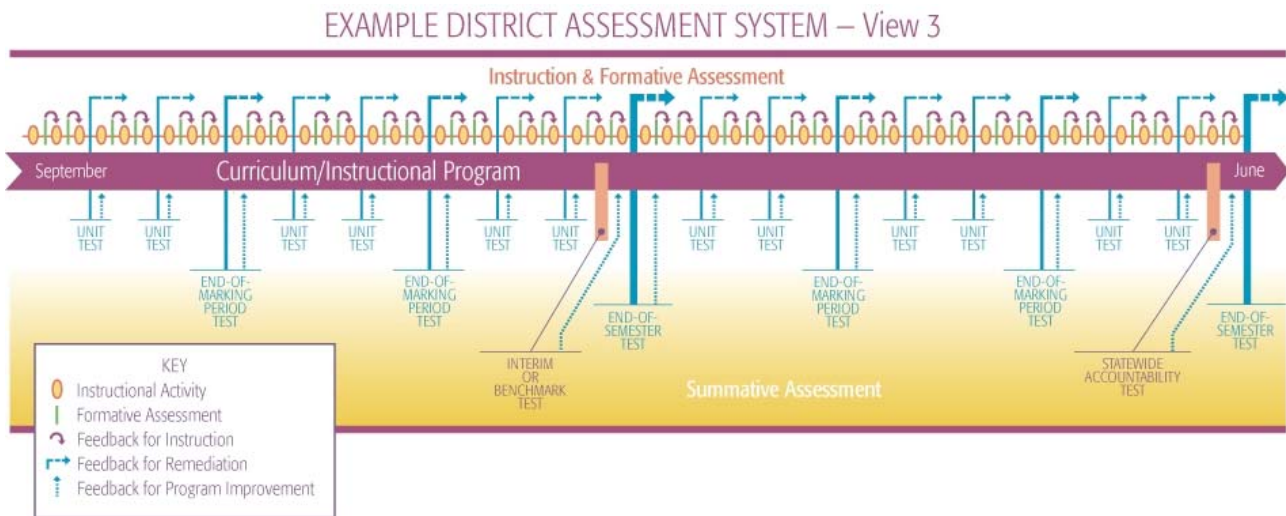
types of information provided by different types of tests.

Exhibit 1: Different types of tests for different purposes and audiences<sup>2</sup>

Decision?	Who?	Helpful Info?	Type of Test?
What comes next in learning?	Students Teacher Parents	Continuous. Individual student progress on standards.	Continuous. Classroom-based Diagnostic.
Are standards mastered? Is the program working?	Teacher teams Principal Curriculum directors	Periodic. Summarize across classes. Which standards are mastered.	Interim. Benchmark. Common across classes.
Are enough students meeting standards?	School/District /Community /State/Federal	Summarize across schools/districts. Percent mastering standard.	Annual accountability assessment. Reliable for purpose.

How these different types of assessment measures might take place over the course of a school year is shown in Exhibit 2.

Exhibit 2: Chronological View of Administration of Comprehensive Assessment Measures<sup>3</sup>



<sup>2</sup> Source of exhibit: Stiggins, R. (2008). Presentation at “Educational Testing in America: State Assessment, Achievement Gaps, National Policy and Innovations.” Sponsored by Educational Testing Service and the College Board, September 7, 2008, Washington, DC.

<sup>3</sup> Source of exhibit: Kahl, S. (2006). Formative Assessments, Summative Assessments, and District Assessment Systems. Presentation at the annual CCSSO National Conference on Large-Scale Assessment.

A balanced assessment system *provides information about all the students* through appropriately designed assessment measures and inclusive policies. That is, a balanced assessment system provides valid information for all students; this may require accommodations or even different measures. Although much recent attention has been paid on developing appropriate assessments for students with disabilities because of the demands of No Child Left Behind, a balanced assessment system extends this principle to interim measures and formative assessment processes as well.

A balanced assessment system *provides information on valued constructs, content, and skills*. Many states' concerns with current assessment systems are that they may need to be supplemented or expanded to capture what is valued. For example within mathematics, both procedural fluency and ability to solve less-structured problems may be valued. This may imply multiple types of measures to adequately assess the construct. Some interests are to include more than reading or English language arts and mathematics for accountability

purposes. Some are interested in assessing different constructs or learning targets, such as college readiness, "soft skills" such as teamwork, personal traits such as perseverance or punctuality, dynamic skills such as speaking, listening, or performing, and so on.

A balanced assessment system *provides not only information about "what is," but also provides information about "what should be done."* This is especially true of the formative assessment processes. Different formative assessment systems take various approaches to how and how much such information is provided. For example, it may be useful to consider various formative assessment approaches as ranging on a continuum, relatively emphasizing assessment or instruction; another continuum is how much the formative assessment approach focuses on expertise in people or expertise in things such as assessment instruments or technology-based tools. (See Exhibit 2). In whatever form provided, a balanced assessment system should include support for using the information appropriately and effectively.

Exhibit 3: Two design dimensions of formative assessment, with illustrative approaches<sup>4</sup>



<sup>4</sup> From Gong, B. (2007). "Choices within Formative Assessment." Paper presented to the FAST SCASS sponsored by CCSSO.

Note that this definition of a balanced assessment system includes but is much more comprehensive than some of the current concerns about current assessment and accountability systems, such as *No Child Left Behind*, that include:

- ⌘ \* Desire to combine multiple measures into overall decisions (e.g., accountability systems' use of multiple indicators, weights, and combining rules)
- ⌘ \* Concerns that accountability systems have distorted desired focus for learning and teaching (e.g., counteract "narrowing of curriculum")

- ⌘ \* Concerns that current assessment and accountability systems do not provide enough and the right information to support needed improvements in student learning and in school capacity (e.g., provide information to inform more local processes)
- ⌘ \* Desire to include different summaries of school performance into accountability, notably status, improvement, growth (including "value-added"), and acceleration. (See Exhibit 4).

### Creating More Balanced Summative Measures

These are all ways primarily to characterize *summative school performance*, although each uses individual student data—from the annual state assessment. A balanced assessment system may incorporate any and all of these four approaches to

summarizing school performance and including it in school accountability, but is much more extensive than different ways to address accountability.

Exhibit 4: Four views of school performance<sup>5</sup>

	Status	Change
Achievement	"Status": How high do students in this school score on state tests?	"Improvement": Is the performance of successive groups increasing from one year to the next?
Effectiveness	"Growth": Are individual students learning as they progress from one grade to the next?	"Acceleration": Is the school becoming more effective or improving more rapidly?

(Carlson, 2001; Gong, 2002)

<sup>5</sup> Drawn from Carlson, D. (2001). Focusing State Educational Accountability Systems: Four methods of judging quality and progress. Paper downloaded 10/23/08 from <http://www.nciea.org/cgi-bin/pubspage.cgi>. This framework was embedded as a consideration in the design of accountability systems by Gong, B. (2002). Designing School Accountability Systems: Towards a framework and process. Washington, DC: CCSSO.

A balanced assessment system *provides information within a time frame to be useful*. This means that a balanced assessment system recognizes the varying time needs for “shorter cycle” and “longer cycle” feedback. The need for timely information is one reason why a state test, traditionally administered once a year, cannot provide information useful to inform regular instruction—even if the information were focused and specific enough, which is nearly

impossible with regular tests that survey broad domains with relatively few assessment items.

Finally, a real balanced assessment system *embodies choices and compromises* that balance the ideal and the practical. For example, in every assessment there are choices made about standardization and flexibility. Every state will need to make choices regarding costs-benefits. And every state will need to choose what to implement sooner and what to implement later.

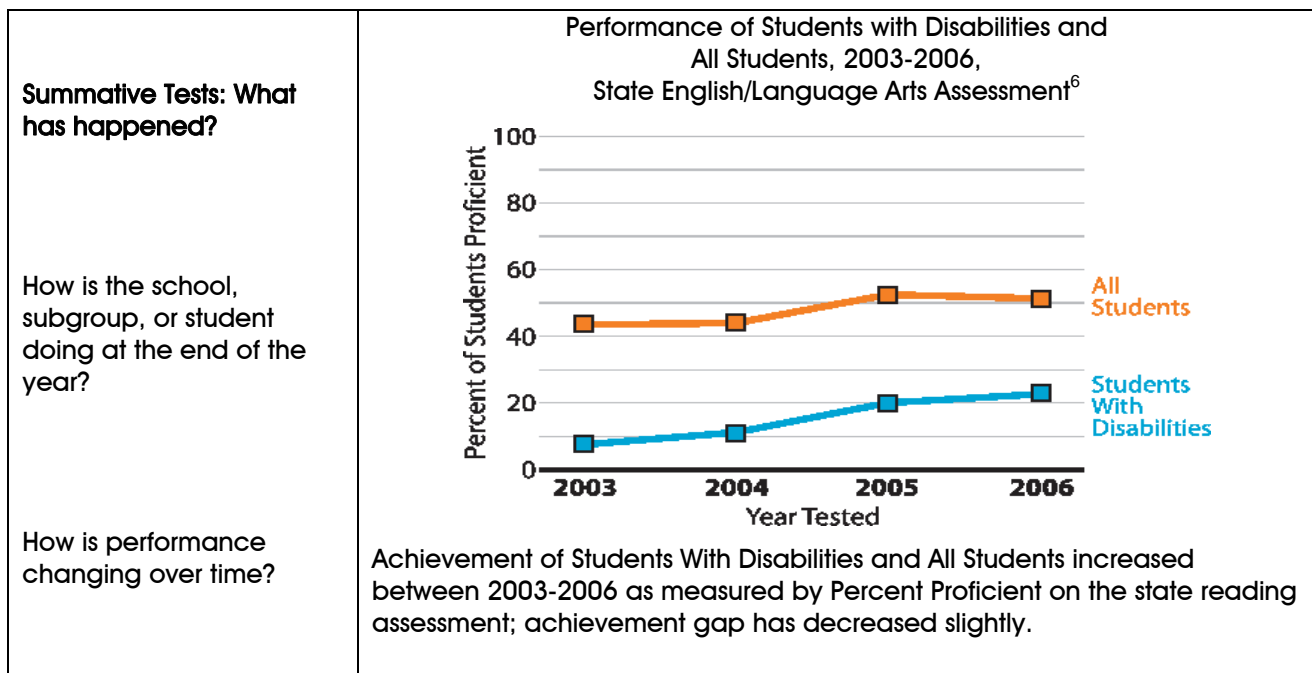
### Why is it important for states to support balanced assessment systems?

It is important for states to support balanced assessment systems because the information provided by state summative assessments are insufficient by themselves to foster the desired student learning and school capacity. The state has a valuable role to play in helping ensure the coherence and equity of balanced assessment systems throughout the state. Balanced assessment systems provide information not only about what has happened, but also about why, and how to improve. Because interim measures and formative assessment processes can be administered more

frequently, they can provide valuable “early warning” information and feedback to help direct program administration, allocate resources, and improve learning and teaching. State tests, by design, do not provide all the information needed to inform program improvement and improve student learning.

The examples in Exhibit 4 illustrate the additional information available through a balanced assessment system that includes different types of assessments at different levels of the educational system.

Exhibit 5: Illustrations of information provided by summative and interim tests, and formative assessment process

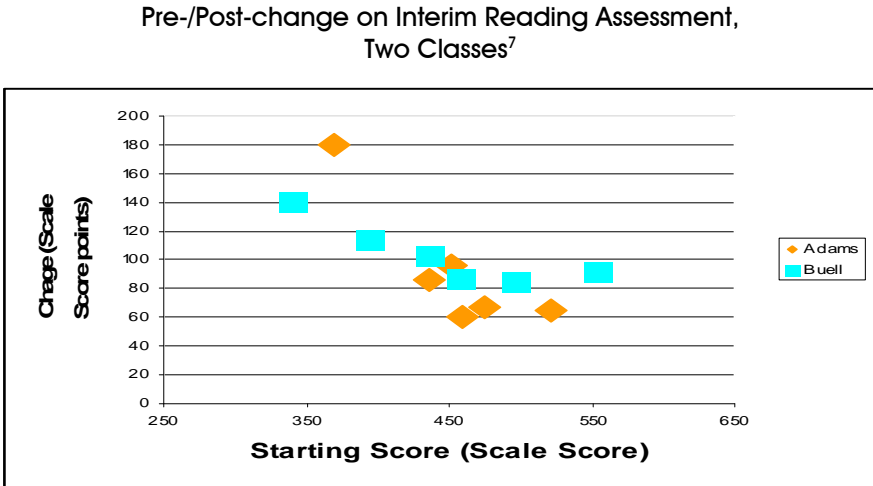


<sup>6</sup> Adapted from Cortiella, C. & Burnette, J. (2008). *Challenging change: How schools and districts are improving the performance of special education students*. New York: NY: National Center for Learning Disabilities, p. 23.

**Interim Measures: What effect has my program had?**

What are the strengths and weaknesses of students or programs during the year?

Why are summative results what they are?



District program to help lower-performing students catch up appears to be working, as measured by gains on interim assessment. Achievement of students who started low (with scale score less than 450) increased relatively rapidly (district goal = at least 80 points). Teachers A and B should discuss what could be done to improve gains of students who are scoring relatively higher (e.g., is a different curriculum needed to challenge these students?)

Formative Assessment process: How can I help this student learn now?

Jennifer has difficulties reading, as evidenced by low scores on interim and classroom assessments. Jennifer’s teacher will work with her to understand and address the causes. Is it due to Jennifer tackling reading material that is too difficult? Is Jennifer using comprehension strategies effectively (e.g., scanning the table of contents prior to reading, or briefly retelling the story to consolidate her understanding)? Does Jennifer know the relevant vocabulary? Is Jennifer able to decode words fluently enough not to interfere with comprehension? Is Jennifer motivated? Jennifer’s teacher will use this information to adjust her teaching and help Jennifer day by day<sup>8</sup>.

<sup>7</sup> Adapted from Growth Report for STAR Reading provided by Renaissance Learning, 2008 (hypothetical data).

<sup>8</sup> Adapted from “Manage Each Student’s Reading Practice” in *Getting Results with Accelerated Reader* (2007). Wisconsin Rapids, WI: Renaissance Learning.



### *How do purpose and use affect the design of an assessment?*

It is a principle of test design that form follows function—that is, the design of the assessment reflects the intended purpose and use.

For example, it is widely agreed by assessment authorities that a traditional state assessment cannot provide useful “diagnostic information” for an individual student. Three reasons are commonly given for this. First, to be diagnostic the information must be precise about what the individual student is stronger and weaker on. Most state tests are designed to survey key topics in a broad array of content standards. They simply are not fine-grained enough to provide much information about any particular bit of knowledge or skill. For example, a typical state test will have 45-60 items. In math, the test typically is targeted at 30 or more content standards or large learning objectives. The test can provide information on the student’s general proficiency, but cannot provide much information about any one topic. Second, the state test is necessarily designed for efficiency the timing is often out of synch. To be useful as a

diagnostic test, the test should provide information close to when useful instructional action can be taken. The results of the state test rarely provide reliable results about sub-skills or subscores, primarily because so few items are available—often only 1-3 for a particular sub-skill that is assessed. In addition, the multiple choice format severely limits what can be inferred reliably about how a student approached a problem or why they got it correct or incorrect. Third, state tests occur once a year, and as such their results usually are not timely to inform learning in close connection with the regular school learning sequence.

Another example is developed below to illustrate this point that different assessments are designed to serve particular functions and provide particular types of information. The example chosen is for interim measures, since the use of interim instruments is rapidly expanding, but without the intense scrutiny that now accompanies the development and implementation of state summative tests.

#### **Design of Interim Assessment Measures: Four Purposes, Four Designs**

Below are four test designs that differ in terms of the content included in a set of interim assessment measures administered four times during the year, followed by the state summative tests. For the purpose of this illustration, assume that the district has established a learning sequence of topics corresponding to state content standards, and organized them into an instructional sequence. The ten topics, A through J, are taught one a month. Some of the topics have multiple parts, such as D<sub>1</sub>, D<sub>2</sub>, D<sub>3</sub>, and D<sub>4</sub>. The state assessment does not include all of the topics taught (for example, it includes standard D<sub>4</sub>, but not standards, A, B, D<sub>1</sub>, etc.

The four interim instrument designs could all be used to predict how well a student might do on the state test, for example. However, the designs have different assumptions (notably about students’ remembering/forgetting over time) and very different score structures. For example, a score of “50% of the items correct” would have very different interpretations of how well a student was prepared, for the various designs, where 50% would be a high performance in the first design (since the student has not yet been instructed on most of the content included on the test), and a low performance in the second design (where ostensibly the content was all taught recently).

Note that in every one of these designs, 100% of the test items are aligned to the state’s content standards (one-way alignment between items and standards).

### Exhibit 6: Design of Interim Assessments

<b>Learning sequence of 10 topics/content standards during year</b>									
<b>A</b>	<b>B</b>	<b>C</b>	<b>D<sub>1234</sub></b>	<b>E</b>	<b>F<sub>123</sub></b>	<b>G</b>	<b>H</b>	<b>I</b>	<b>J</b>
Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	June

<b>Four interim assessment instruments &amp; content topics assessed</b>	<b>State test &amp; content assessed</b>
--	--

C, D4, F <sub>2</sub> , etc.	C, D4, F <sub>2</sub> , etc.	C, D4, F <sub>2</sub> , etc.	C, D4, F <sub>2</sub> , etc.	C, D4, F <sub>2</sub> , etc.
---------------------------------	---------------------------------	---------------------------------	---------------------------------	---------------------------------

In this model, the interim assessment instruments mirror the end-of-year state test in terms of content, balance of emphasis, format, administration conditions, etc. Each test administered during the year covers the same content and has the same design. This design provides high “practice” and high “prediction” from the interim to the end-of-year state test. It is also an excellent design for program evaluation of the impact on learning of an instructional program between pre- and post-tests.

A, B	C, D	E, F	G, H	C, D4, F <sub>2</sub> , etc.
------	------	------	------	---------------------------------

In this model, the interim assessment instruments focus on the content that was instructed. Each interim measure covers only the content in the most recent instructional period, and thus each test’s content differs from the others. This may be the best design for assessing recent instruction and informing remedial work on what was recently instructed. It may not be an effective predictor of student performance on the state test if students forget after instruction.

A, B	A, B, C, D	A, B, C, D, E, F	A, B, C, D, E, F	C, D4, F <sub>2</sub> , etc.
------	---------------	---------------------	---------------------	---------------------------------

In this model, the interim assessment instruments are designed to assesses what was instructed, but is cumulative, i.e., the assessment includes all topics instructed up to that point in time. This model values student retention of knowledge previously taught. It may not be an effective or efficient way to predict student performance on the state test.

A, B	B, C, D	C, D, E, F	C, D, F <sub>2</sub> , G, H	C, D4, F <sub>2</sub> , etc.
------	---------	---------------	-----------------------------------	---------------------------------

In this model, the interim assessment instruments are designed to assess what was instructed, but are also cumulative for the topics that will be assessed on the state test.



## What is the relation between assessment, curriculum, and instruction in a balanced assessment system?

When there is a strong model of curriculum and instruction, the assessment can be both more focused and less extensive. It can be more focused because it can be designed to match known learning targets, with known parameters such as dependencies, timing, and customization for certain students. Importantly, when the assessment is matched with a known curriculum or instructional model, the assessment can be more focused and powerful because there are defined user models and supports.

Summative tests are usually intended to check what students know and can do that has been acquired over longer periods of time—a semester, a year, or even multiple years. These instruments provide valuable information on what students remember and what they can apply to problems and situations not related necessarily to instruction that immediately took place. An important aspect of interpreting such test results is that these typically represent what the student knows and can do, and that performance is independent of particular instruction or a particular curriculum. Most summative tests in fact are designed not to be sensitive to any particular curriculum or means of teaching.

Interim assessment measures may be more closely specified for specific content, curricula, and instructional methods. This is particularly true when the interim instruments are developed by those who know the curriculum, whether local education agencies or commercial companies that are marketing curriculum and assessment materials.

Interim assessment measures, however, may be designed to be less closely aligned with particular curricula, and be more like summative tests. In fact, some interim instruments are designed to predict performance on the summative test, and so replicate the summative assessments in design as much as possible. Such “early warning” instruments are used primarily to let teachers, administrators, and students know which students are likely to pass the state test, and which are likely

to need additional supports. Such interim assessment measures, like summative tests, provide only broad guidance on student strengths and weaknesses, and so must usually be supplemented by more fine-grained instruments in order to diagnose what exactly a student may need help with.

Such diagnosis is integrated with instruction and learning in a dynamic cycle where assessment measures portray student achievement (and strengths and weaknesses) and are used to inform instruction, and where instruction learns from assessment information and is modified to better achieve the learning goals.

The pattern of using assessment to guide instruction and of instructional goals guiding assessment is a powerful symbiosis. The key elements are assessment measures that yield relevant information to guide instructional actions, and instruction that is modified in response to assessment and other data in an on-going basis. These key elements are found in a number of approaches now, including “data-driven decision-making,” “computer-based management,” “response to intervention,” and “formative instruction.”

The integration of assessment and instruction has long been a goal of those who have envisioned powerful learning at the classroom level. One way assessment and instruction have been merged is through infusing activities into the curriculum that are designed to provide assessment opportunities. There are many ways this can be done. For example, a teacher might embed a pre-assessment into the schedule a week prior to beginning a unit. The pre-assessment is a learning activity that serves as an advance organizer for the students. It also provides the teacher formative assessment information so she can pre-teach students who may be weaker, and plan for how to capitalize on the strengths of students who are already well-developed in some knowledge and skills in that unit. A teacher might also integrate assessment and

instruction by regularly providing activities that involve students in meaningful practice and feedback. Homework and worksheets are typical ways this is done. However, to be effective at increasing student competence, the practice must be well-designed and implemented. Research supports the teacher who observed, “Practice doesn’t make perfect. Good practice with effective feedback makes perfect.” One critical aspect of “good practice” is practicing the right things, or not reinforcing the wrong things. Two critical aspects of effective feedback is that it is timely, and that it leads the student to know what to do better, not just what was done wrong.

Having regular routines of seeking data and modifying actions is a hallmark of higher-performing organizations. Balanced assessment systems are most effective when they combine the assessment elements—so that assessment data are gathered—and also routines or processes that make the use of the assessment data focused and coherent. Such

processes often need to be in place for several cycles in order to be tuned to be effective and to allow the

persons—superintendent, principal, teachers, parents, policymakers—to develop skill in interpreting and using the data.

These routines of how to use specific data may be developed by the organization. Some commercial assessment products include training of how to not only interpret the scores, but of how to use them. A few commercial assessment products include multiple elements of a balanced assessment system—for example, different assessment material useful at the daily, weekly, monthly, and longer time spans—and provide specific guidance and training on how the different measures should be used in a coherent way. Most states have suggested ways to use data, but states’ training typically focus on using data from state tests, and so often are limited in the scope that a balanced and comprehensive assessment system would

**Effective practice**, according to one source<sup>1</sup>, involves six characteristics:

- ∞ Dedicate enough time to practice – most things worth learning cannot be learned with one exposure or in a few minutes. Many skills require long periods of practice over many years.
- ∞ Practice the right things – and often what one student needs to work on will be different from what another student needs.
- ∞ Work until the desired level of mastery is reached. That may involve fluency as well as accuracy, flexibility as well as generalization. Formative assessment should inform work in a “zone of proximal development” that allows reasonable work-to-mastery with goals set iteratively higher.
- ∞ Practice is informed and adjusted according to assessment feedback, and vice versa.
- ∞ Assessment, instruction, and feedback should engage students and build their confidence. Certainly student affect is recognized as important in contributing to student learning and student self-assessment.
- ∞ Practice should include review to consolidate knowledge and skills. Consolidation should be incorporated into assessment, practice, and instruction over varying time spans, depending on the knowledge and skills and the learning goal.

1 “Getting Results with Accelerated Math”. Renaissance Learning, (2008), pp. 6-7.

ideally cover.

## Exhibit 7: Student Proficiency Case Scenario

### A Student Use Case Scenario: Helping Devon learn to proficiency<sup>1</sup>

*This scenario illustrates the coherent use of a balanced assessment system to help an individual student achieve his learning goals. We refer to it as a “use case scenario” because it illustrates not only the features of multiple types of assessment—summative, interim, and formative—but also how they are used individually and in relation to each other.*

Devon was a low income, non-native English speaking 6<sup>th</sup> grader in a middle size school district in the 2007-2008 school year.

That year Devon took all state tests during the spring testing window. His scores showed that he was below proficiency in math and just above proficiency in reading. He will move from the district elementary school to the middle school in the fall. School officials are concerned about increasing Devon’s academic achievement and managing the transition from elementary to middle school.

This is a timeline of how Devon and his teachers might interact in a balanced and coherent assessment system:

**Summative** – Devon takes state summative test in spring. Teachers and administrators consider results for Devon in an overall learning plan that includes academic and non-academic considerations.

**Interim** - Early in the Fall Devon takes the districts interim instrument that give teachers an insight about where he might score on the state tests in the spring and what areas are most at risk. His scores indicate that he is likely to score below proficiency in both reading and math.

**Formative** – The reading and math teachers meet with the ESOL resource person and the curriculum facilitator. They decide to test Devon over specific content objectives to determine more specifically where his deficits are in order to decide on an instructional plan. The reading diagnostic indicates that Devon has a problem with the necessary vocabulary. They decide to leave him in the general reading class with additional attention to problem vocabulary. In math, however, there are multiple areas where Devon is at least a grade behind. The teachers decide to add him to an additional class taught by a math resource teacher and monitor his progress with a focused assessment that can be used weekly.

**Formative** – Teacher observation, class grades, and weekly assessment scores indicate Devon is having some trouble managing the transition and learning the math curriculum. During parent-teacher conferences, it is decided to have Devon attend an after-school math tutoring session.

**Interim** – The second interim instrument indicated Devon is on trend to meet proficiency in reading, but, although making progress on some objectives, is still at-risk in math.

**Formative** – Teachers decide to maintain the reading intervention and re-focus the instruction and continue monitoring progress.

**Formative** – Teacher observation and weekly scores indicate Devon is making progress.

**Interim** – The third interim instrument indicates Devon is on trend in reading but still below in math.

**Formative** – Teachers decide to focus on a smaller number of key math objectives and coordinate this decision with the regular math class, the math lab, and the after-school tutor.

**Summative** – Devon takes the state reading and math assessments in the spring. He is proficient in both reading and math.

**Formative** – Teachers analyze Devon’s scores and progress on all measures, form a plan for the rest of the year, end the after school tutoring session, and prepare of report for the 8<sup>th</sup> grade team.

This timeline summarizes how Devon’s teachers used summative, interim and formative assessment interactively in a balanced, coherent, and systematic way to help support his learning goals.

Summative – Devon is at risk in math  
Interim - Devon is at-risk in math  
Formative – Intervention is planned  
Formative – Progress monitoring by teacher  
Interim – Some progress  
Formative – Intervention continues  
Formative – Additional strategies are utilized  
Interim – Progress is better  
Formative – Intervention is adjusted  
Formative - progress monitoring by teacher  
Summative – Devon takes state NCLB assessments  
Formative – Progress is analyzed

<sup>1</sup> – This case scenario provided by Tom Foster, Kansas Department of Education

## Exhibit 8: District-School-Classroom Case Scenario

### A District-School-Classroom Use Case Scenario

*This scenario focuses on how a district superintendent, school principals, team lead teachers, and individual teachers are organized to make comprehensive assessment systems informative to support student learning.*

This scenario reflects the comprehensive assessment pattern shown in Exhibit 2. In addition, it reflects recognition that several students in the district needed not just differentiated instruction, but tiered intervention. (See Exhibit 9).

Teachers worked individually and as grade-level teams to diagnose student needs and match them to an appropriate instructional program. The district superintendent works with the building principals to ensure resources are focused on helping children achieve challenging learning goals. For example, students who are identified as at risk of not scoring proficient on the state summative test by the end of the year are provided extra support. Principals, counselors, and lead teachers disaggregate the state test results to identify sub-areas where students are likely below proficient or just slightly above proficient.

They administer an interim instrument to get additional diagnostic and predictive information, and combine with teacher information to confirm their diagnoses.

Students who are identified as having major weaknesses in a particular area are placed in a supplemental 6-week course that focuses on that area. The supplemental course uses an intensive formative assessment process coupled with specific instruction to help students catch up in that area. The principal, counselors, and teachers attend to motivational effects so that students can move out of this remedial course quickly. The principal tracks the progress of students through use of “dashboard software” that provides up-to-the-minute summaries of performance on the assessments for individual students, classes of students, grades, and the school as a whole.

Teachers meet in grade-level teams weekly to discuss how to adjust their lessons. Their discussions are informed by the common assessments students take weekly (quizzes), as well as individualized assessments and the classroom assessment information gathered by the teachers.

Individual teachers work with students on individualized learning plans that combine group and individual instruction, practice, and assessment.

## Exhibit 9: Tiered Levels of Intervention

Tiered Levels of Intervention (adapted from Daro, 2008)			
Situation of Student	Needed by Student	Intervention Tier	Intervention
Keeps up	Regular instruction	NA	None
Struggles some assignments	Extra feedback on work, thinking; focused practice	Tier 1	Classroom formative assessment (e.g., teacher Q&A, peer assessment); effective practice <i>Professional development</i>
Not bringing enough from earlier lessons each day	Extra support with regular program		Homework clinic, tutoring, guided practice, attention beyond regular class <i>Scheduling / targeted use of adopted materials</i>
Misconceptions disrupt participation and success in mathematics (gaps)	In depth concentration on troublesome concepts (not initial teaching)	Tier 2	Sustained instruction with special materials beyond regular class period and/or summer school
More than a year behind; misconceptions or gaps from many years	Extra time and focus on critical mathematics to accelerate to grade level	Tier 3	Designed double period ramp-up course, summer school
Source: Adapted from Daro, P. (2008). Presentation at the Math Symposium sponsored by the California Association for Supervision and Curriculum Development, Feb. 5, 2008. Retrieved 2/6/09 from the web at: <a href="http://www.cascd.org/math_symposium/Math%20Symposium%20-%20Daro.pdf">http://www.cascd.org/math_symposium/Math%20Symposium%20-%20Daro.pdf</a>			

## How can states support the development and implementation of balanced assessment systems?

States can play a pivotal role in supporting the development and use of balanced assessment systems. They can:

- Ensure the integrity of the state-level components and the usefulness of the summative test data.
- Foster balanced assessment systems through appropriate policy. The policy should establish the need for balanced assessment systems, promote their development, remove barriers, and support professional development and infrastructure.
- Help ensure the coherence and equity of balanced assessment systems throughout the state.
- Directly develop some elements of a balanced assessment system, particularly interim instruments. However, unless the state provides a detailed curriculum, it must support the local control over choice of interim instruments and formative assessment process and how those data are used. Locally developed and commercially available instruments hold great promise but must be carefully reviewed and used appropriately.

## How can states support the development and implementation of balanced assessment systems?

States can play a pivotal role in supporting the development and use of balanced assessment systems in several ways. This report discusses several key ways below. Similar advice could be given about the role that district and school administrations can play as well.

States must *ensure the integrity of the state-level components*. First, states must be sure that the state content standards that define what the state expects students to know and be able to do are clear, coherent, feasible, and communicated appropriately. Second, the state should ensure its state tests are valid, reliable, and useful. That involves ensuring the instruments are designed to assess the content standards well, and that the scores meet psychometric standards of quality. Assessment results must be provided in ways that are accessible to parents, teachers, school administrators, policymakers, and other key users. Often training and professional development are necessary to ensure assessment results are used to their full potential and are not misinterpreted or misused.

States can *create a policy environment that fosters balanced assessment systems*. That policy environment will include communicating the need for balanced assessment systems. It will also remove barriers to the establishment of balanced assessment systems, such as the widespread notion that practicing on released items from the state assessment is a sufficient instructional strategy. Policy support should also include funding and attention to other resources, such as professional development. A key role state's policies should address is the coherence of balanced assessment systems. In every state, districts and schools are establishing the interim instruments and formative assessment process, the curriculum and instruction, and the structures, policies, and programs of how assessment information will be used. Often these involve resources developed by the district, school, or classroom teacher, or purchased or developed in conjunction with a commercial vendor or other partner. State policies must consider the wide range of possible ways to construct a balanced assessment system.

States can also *work directly to establish the components of a balanced assessment system*. Many states are supporting the development and use of interim and formative assessments in many ways. States are using different strategies, depending on the philosophy, capacity, and resources of the state and the local districts and schools. Some states are developing interim instruments or a secure item bank that can be used by districts. A few states are reviewing and developing “approved lists” of interim assessment products and providers. Most states have relied on local judgment and initiative to develop balanced assessment systems at the district, school, and classroom levels. Many states encourage the use of interim instruments by providing some funding and evaluation tools that districts and schools may apply themselves as they choose to purchase or develop their own interim instruments. True formative assessments that are closely

integrated with minute-by-minute and day-by-day instructional decisions are typically developed locally, although there are some commercial or free assessments that can be quite useful if used appropriately. All states recognize the need for more training and professional development in selecting, using, and interpreting the results of interim and formative assessments but most states acknowledge they could do more to provide more funding and direction for professional development regarding balanced assessment development and use.

States have taken various approaches to supporting the development of balanced and coherent assessment systems. States’ approaches may be characterized in how large a role the state takes, or how localized the balanced assessment system is. (See box on next page.)

A state supports the development and use of balanced assessment systems by:

- Ensuring the clarity, coherence, and usefulness of the state’s content standards, communication, and professional development regarding the state’s summative tests.
- Providing guidance to local districts and schools regarding the selection and use of interim assessment. Some states may provide interim assessment instruments; other states may provide guidance or other support where districts and schools make their own decisions about adopting interim assessment measures.
- Providing strong policy support and training for effective formative assessment processes.



## *Different State Approaches to Developing Balanced Assessment Systems*

### Highly Centralized

- ↻ Develop and promulgate tight definition of balanced assessment systems and components (e.g., interim and formative assessment)
- ↻ Define a state curriculum
- ↻ Establish state criteria and vetting process for interim assessment measures
- ↻ Develop interim instruments and/or item banks and supporting software for local use as interim or in a formative assessment process
- ↻ Require specific formative and interim assessment practices by schools identified through accountability system
- ↻ Tie state funds to local adoption of highly specified assessment uses
- ↻ State offers specific professional development around use of assessment information

### Moderately Centralized

- ↻ Define a “model curriculum” but expect much local variation of learning targets, timing, and standardization
- ↻ Make state test data available in machine-readable formats (via state data warehouses or downloaded/CDs) so local education agencies can integrate with local assessment and other data
- ↻ State provides infrastructure for local development of balanced assessment system components (e.g., develops data warehouse and allows upload of local assessment information and local custom reporting; develops templates, guides, and other tools for interim assessment alignment studies)
- ↻ State develops list of recognized vendors or establishes selective partnerships with professional development/technical assistance providers around assessment
- ↻ State develops models of balanced assessment components and/or processes
- ↻ State provides quality assurance and technical support for local implementation of balanced assessment components (e.g., formative assessment)
- ↻ State provides some financial support and/or incentives to develop and use a balanced assessment system

### Highly Localized

- ↻ Promote highly individualized learning targets
- ↻ Provide funds for local education agencies to engage in assessment professional development selected by LEA
- ↻ Fuel attention to balanced assessment systems and the underlying notions of educational reform; provide a state “theory of action” for educational reform that includes balanced assessment systems

## Developing Comprehensive Assessment Systems

### *Can a state build local or interim assessments from the state summative tests?*

Several states have developed interim assessments for use by districts, schools, and teachers. These interim assessments typically are used to:

- Predict student performance on the “real” state test, presumably with the intention that students who score low will be provided with additional support to help them score better.
- Provide practice and familiarize students with the format, instructions, and general nature of the state summative test. The assumption is that students should be familiar with these aspects of the test in order to be able to do their best and get a valid score.
- The state may provide these interim assessments in downloadable form, printable files from the web or as downloaded. Some states also provide an on-line version which allows on-line scoring and reporting. The most sophisticated systems may offer computer-adaptive administration, whereby the computer selects an item from the item bank based on whether the student answered the previous question correctly.
- Some districts may use these state-provided interim instruments for program evaluation to identify areas of strengths and weaknesses in the curriculum or performance of particular teachers, departments, and/or schools.

Some states have also built collections of assessment items, called item banks. Item banks typically allow a user, such as a teacher or district staff person, to draw upon items when constructing a test. Item banks provided by the state typically are used to:

- Provide items with ostensibly better quality than items from other sources. States are assumed to have reviewed the items in their item banks to ensure that each item: 1) is aligned to the state content standards, 2) has undergone appropriate reviews for sensitivity

and bias, 3) is accurately described and coded, and, 4) is free from copyright and other restrictions for use by teachers and other authorized users. More complete items banks also provide statistical information, such as IRT difficulty parameters.

- Several states include software to help teachers and other use the item bank. The software might include facilities to assemble the items into a test, print, and create supporting materials such as answer key, content standards codes, etc.

### *Can a state build a state accountability system from local assessments?*

- It has been challenging for local assessments to be used in state accountability systems. There have been few attempts, and even fewer efforts that have lasted several years.
- The most notable use of local assessments is the Kentucky writing portfolios. Kentucky writing portfolios consist of writing produced within the classroom setting and scored locally. The scores are included as part of the writing scores used to hold schools accountable in the state accountability system. The writing portfolios have been a part of the Kentucky assessment and accountability systems since 1991.
- The Kentucky state assessment included performance tasks from 1991-1998. The Kentucky performance tasks were characterized as having a component done as a group and a component done by individual students; often involved hands-on manipulative materials; took about an hour to administer including students writing their responses. Performance tasks were used in Kentucky to assess a variety of content areas including mathematics, science, social studies, arts and humanities, and vocational studies/practical living. The performance tasks were dropped from use in the large-scale assessment due to concerns about equating and cost.

- Nebraska helped districts implement local assessment systems, and used the results for its state assessment, including for NCLB. Nebraska did administer a common writing assessment statewide, and conducted expert and peer reviews of the local assessment systems and assessment instruments. Nebraska phased out from using local assessments starting in 2008 in compliance with a state law that required common state assessments in ELA, math, and science. The change was perceived by many to be due in large part to Nebraska not being able to get their assessment system approved by the U.S. Department of Education for NCLB.
- States have developed performance assessments for students with severe cognitive disabilities who are not able to validly participate in the regular state assessments even with all possible accommodations. NCLB requires states to provide such assessments, which are commonly referred to as “1%” assessments because of the NCLB stipulation that no more than 1% of the total student population may be declared proficient for purposes of NCLB school accountability on the basis of these assessments’ results. States’ designs for these “1%” assessments vary, but many states use portfolios, observational checklists, and other forms of local assessment. Alternate assessments of alternate achievement standards (AA-AAS) are almost locally scored in the large majority of state programs.
- Several states have required local assessments be used for student accountability, especially exit from high school. For example, Oregon and Washington require students to generate exhibition projects or work samples in order to graduate. The projects are developed in the classroom and scored locally.
- Wyoming has developed a system to support students’ development of a “body of evidence” that the students have met state proficiency standards. The evidence is based on locally administered assessments. Many of the assessments are essay tasks embedded in curriculum units that are used state-wide.
- Rhode Island is developing a system that requires districts to assess students on the basis of local assessment information in conjunction with the state high school assessment in order to make a determination whether the student met locally determined “proficiency-based graduation requirements.” The state has implemented a process for reviewing the districts’ assessment and support systems.
- Many districts have developed assessment and accountability systems that do not use state summative test information. Often these systems use a combination of locally developed and commercially obtained assessment products.

## Where can I go for more information?

CCSSO is a good resource for states to keep up with the rapidly evolving documentation of state strategies and policies around balanced assessment systems. See [www.ccsso.org](http://www.ccsso.org).

1. A strong rationale for a balanced assessment system is given in Rick Stiggins’ (2008) *Assessment Manifesto: A call for the development of balanced assessment systems*. Portland, OR: ETS Assessment Training Institute. The paper also provides an excellent short list of references, especially regarding formative assessment (“assessment for learning”).

2. A good description of interim assessments is provided in *The Role of Interim Assessments in a Comprehensive Assessment System: A policy brief* authored by Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007), and issued jointly by The Aspen Institute, the Center for Assessment, and Achieve.

3. CCSSO is a good source for the evolving agreement about formative assessment and supporting elements, such as learning progressions.

4. More specific information may be available from your state department of education.

5. Several states who participated in this project are excellent sources for information. The following websites are examples of places to start.

AR <http://arkansased.org/>

FL <http://www.fldoe.org/>

GA <http://www.doe.k12.ga.us/>

ID <http://www.sde.idaho.gov/>

KS <http://www.ksde.org/>

MO <http://dese.mo.gov/>

MS <http://www.mde.k12.ms.us/>

NC <http://www.dpi.state.nc.us/>

OH <http://www.ode.state.oh.us>

WI <http://dpi.wi.gov/>

WV <http://wvde.state.wv.us/>

---

### *Endnotes*

---

1 This section is quoted with minor edits from Philip, F. (2008). "What Do We Mean by Coherent and Balanced Assessment Systems"? Washington, DC: CCSSO.

2 Source of exhibit: Stiggins, R. (2008). Presentation at "Educational Testing in America: State Assessment, Achievement Gaps, National Policy and Innovations." Sponsored by Educational Testing Service and the College Board, September 7, 2008, Washington, DC.

3 From Gong, B. (2007). "Choices within Formative Assessment" Paper presented to the FAST SCASS sponsored by CCSSO.

4 Drawn from Carlson, D. (2001). Focusing State Educational Accountability Systems: Four methods of judging quality and progress. Paper downloaded 10/23/08 from <http://www.nciea.org/cgi-bin/pubspage.cgi>. This framework was embedded as a consideration in the design of accountability systems by Gong, B. (2002). Designing School Accountability Systems: Towards a framework and process. Washington, DC: CCSSO.

5 Adapted from Cortiella, C. & Burnette, J. (2008). Challenging change: How schools and districts are improving the performance of special education students. New York: NY: National Center for Learning Disabilities, p. 23.

6 Adapted from Growth Report for STAR Reading provided by Renaissance Learning, 2008 (hypothetical data).

7 Adapted from "Manage Each Student's Reading Practice" in Getting Results with Accelerated Reader (2007). Wisconsin Rapids, WI: Renaissance Learning.